

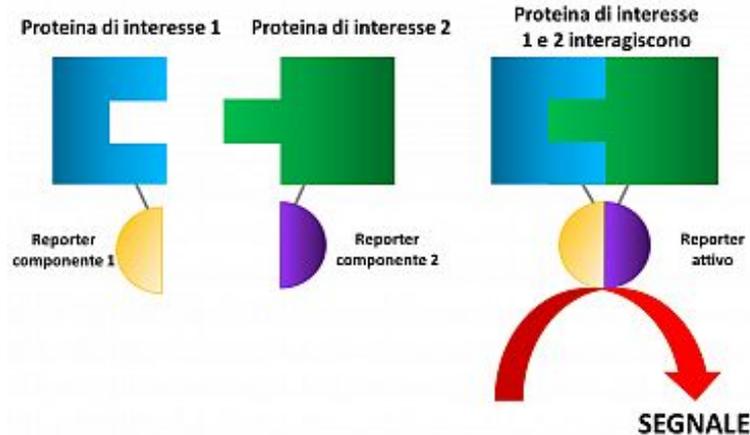
PROTEINE

- Annotazione funzionale proteine



L'annotazione funzionale delle proteine è importante perché ci consente di comprendere come le proteine svolgono il loro ruolo all'interno di un organismo e come esse interagiscono con altre proteine e molecole all'interno della cellula.

Inoltre, l'annotazione funzionale delle proteine è cruciale per comprendere le basi molecolari di molte malattie umane, per lo sviluppo di nuove terapie farmacologiche, per lo studio della biodiversità.



PROTEINE

- Annotazione funzionale proteine

Valutare l'**affidabilità** delle annotazioni è un passaggio critico nell'analisi delle proteine, in quanto garantisce la qualità delle informazioni ricavate e l'accuratezza delle interpretazioni e delle conclusioni tratte.



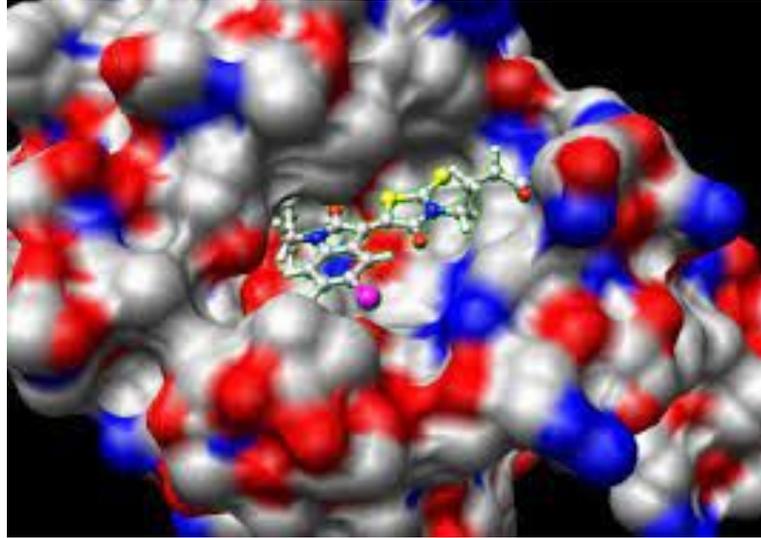
Questo è particolarmente importante in quanto le annotazioni sono spesso basate su predizioni o inferenze, piuttosto che su dati sperimentali diretti.

➔ L'utilizzo di metodi statistici per valutare l'affidabilità delle annotazioni ci permette di valutare la probabilità che un'annotazione sia corretta, in base ai dati disponibili e all'accuratezza dei metodi utilizzati.

PROTEINE

- Interazioni proteiche

Per studiare le interazioni proteiche, i ricercatori utilizzano spesso il **docking** proteico, una tecnica di modellizzazione molecolare che mira a predire la struttura tridimensionale di un complesso proteico formato da due o più proteine.



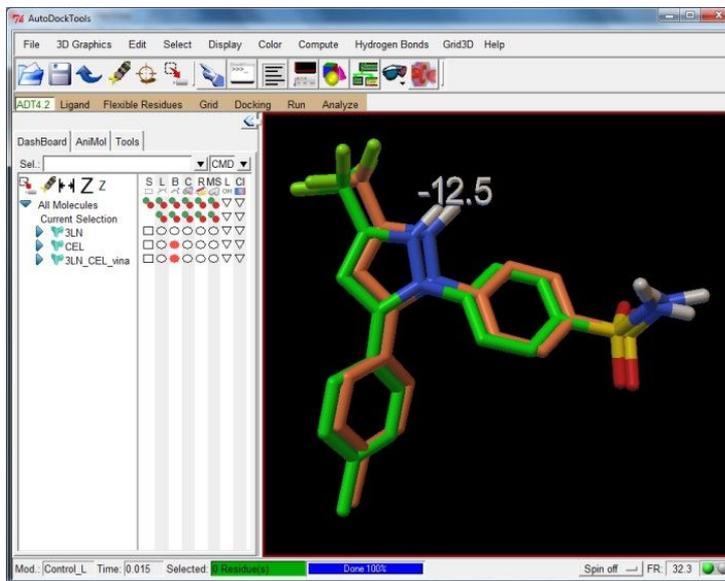
PROTEINE

- Interazioni proteiche



Il docking proteico si basa sull'interazione tra i siti di legame delle proteine e tiene conto delle informazioni sulla **conformazione** e sulla **flessibilità** delle molecole.

Una volta che abbiamo identificato le interazioni proteiche, possiamo utilizzarle per comprendere meglio la **funzione** delle proteine coinvolte e per studiare i meccanismi che regolano i processi biologici.



PROTEINE

- Interazioni proteiche

In questo contesto, le banche dati proteiche giocano un ruolo fondamentale. Le banche dati proteiche contengono informazioni sulle proteine e sulle loro funzioni, tra cui le interazioni proteiche note.



Queste banche dati consentono ai ricercatori di accedere a informazioni precise e aggiornate sulle proteine di loro interesse e di utilizzarle per analisi di tipo bioinformatico, come la predizione di nuove interazioni proteiche o la caratterizzazione di interazioni proteiche note.



PROTEINE

- Banche dati proteiche

Uniprot è una delle più grandi banche dati proteiche al mondo, gestita congiuntamente dall'European Bioinformatics Institute (EMBL-EBI), dal Protein Information Resource (PIR) e dal Swiss Institute of Bioinformatics (SIB).



PROTEINE

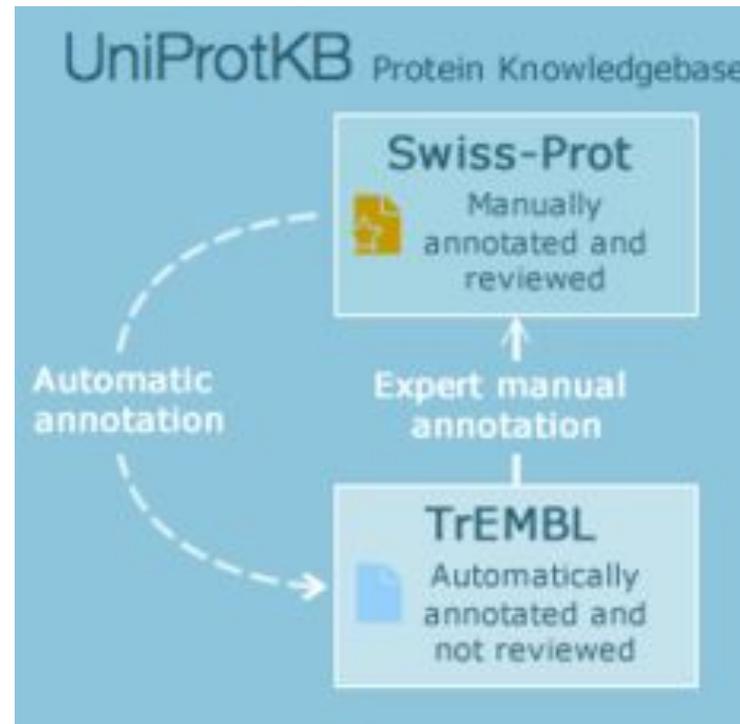
- Banche dati proteiche

Uniprot è formata da due componenti principali: Swiss-Prot e TrEMBL.



Swiss-Prot è una sezione di Uniprot che contiene proteine annotate **manualmente** e curate, con informazioni accurate e complete sulla funzione, la struttura, l'interazione e la localizzazione subcellulare delle proteine.

TrEMBL, d'altra parte, è una sezione di Uniprot che contiene proteine annotate **automaticamente** tramite software di annotazione, principalmente basate su predizioni bioinformatiche e omologia con proteine conosciute. Queste annotazioni automatiche sono poi curate e integrate con altre informazioni dalla comunità scientifica.



PROTEINE

- Banche dati proteiche

Cura manuale

La cura manuale consiste in una revisione critica dei dati sperimentali e previsti per ciascuna proteina e anche di ciascuna sequenza proteica stessa.

I metodi di cura applicati includono:

- valutazione di ciascuna sequenza proteica inclusi siti di splicing e siti di scissione post-traduzionale.
- estrazione manuale e strutturazione delle informazioni dalla letteratura
- verifica manuale dei risultati delle analisi computazionali
- estrazione e integrazione di insiemi di dati su larga scala
- aggiornamento continuo man mano che si rendono disponibili nuove informazioni



PROTEINE

- Banche dati proteiche

Annotazione automatica

UniProt ha sviluppato due sistemi di previsione per annotare automaticamente UniProtKB/TrEMBL in modo scalabile con un alto grado di accuratezza:

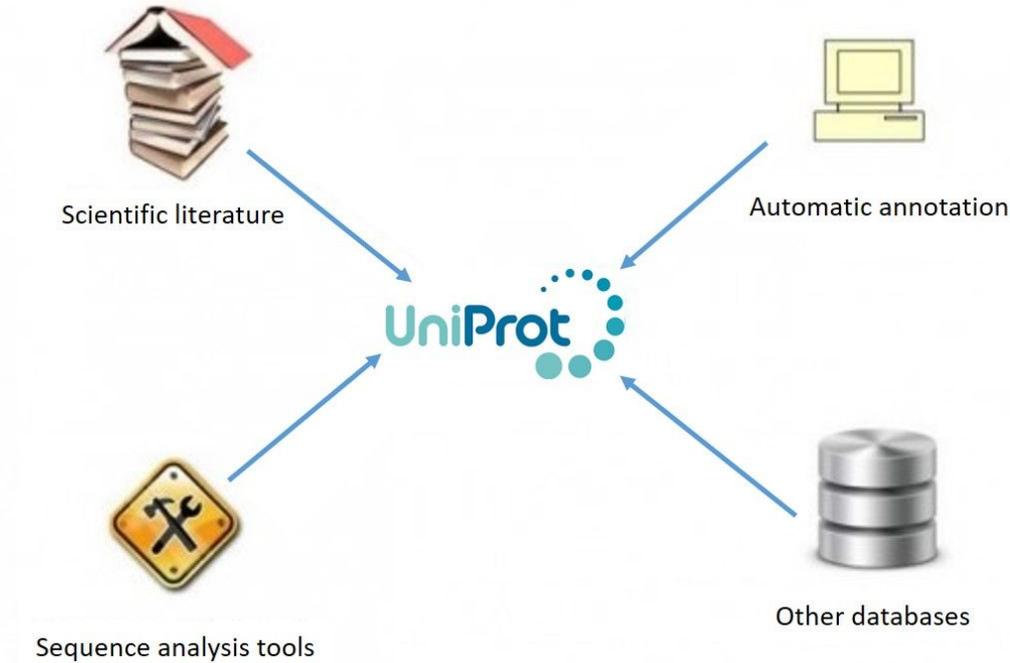
- UniRule è una raccolta di regole di annotazione curate manualmente che definiscono annotazioni che possono essere propagate da voci riviste a voci non riviste in base a condizioni specifiche
- L'Association-Rule-Based Annotator (ARBA) è un sistema di apprendimento multiclasse addestrato su voci annotate da esperti in UniProtKB/Swiss-Prot. ARBA utilizza tecniche di estrazione delle regole per generare modelli di annotazione concisi basati sulle proprietà dell'appartenenza al gruppo InterPro e sulla tassonomia.



PROTEINE

- Banche dati proteiche

UniProt fornisce sia dati di sequenza che informazioni funzionali associate, derivate da una serie di fonti



PROTEINE

- Banche dati proteiche

Le prove utilizzate nelle **asserzioni manuali** sono di colore oro, mentre quelle utilizzate nelle **asserzioni automatiche** sono di colore argento. Le asserzioni manuali includono anche previsioni computazionali o trasferimenti di annotazioni che sono stati successivamente verificati da un biocuratore.



▶ Region	28-123	GFLD subdomain	1 Automatic Annotation	BLAST
▶ Domain	28-189	E1	1 Automatic Annotation	BLAST
▶ Region	96-110	Heparin-binding	1 Publication	BLAST
▶ Region	131-189	CuBD subdomain	1 Automatic Annotation	BLAST
▶ Region	181-188	Zinc-binding		BLAST
▶ Region	194-284	Disordered	1 Automatic Annotation	BLAST
▶ Compositional bias	195-210	Acidic residues	1 Automatic Annotation	BLAST

PROTEINE

- Banche dati proteiche



Alcuni esempi di prove

1. Dati sperimentali: Se un'informazione è stata dimostrata sperimentalmente in un documento, questo sarà indicato con i dettagli del documento utilizzato

Functionⁱ

Catalyzes the addition of fucose in alpha 1-3 linkage. Unlike fut-1, does not add fucose to Man-alpha-1->3-(Man-alpha-1->6)-Man-beta-1->4-GlcNAc-beta-1->4-(Man-alpha-1->6)-Man-beta-1->4-GlcNAc-beta-1->4-(Fuc-alpha-1->6)-GlcNAc-beta-1-Asn (M3F6) or GlcNAc-beta-1->2-Man-alpha-1->3-(GlcNAc-beta-1->2-1->4-(Fuc-alpha-1->6)-GlcNAc-beta-1-Asn (GnM3F6) acceptors. 1 Publication

Catalytic Activity

a beta-D-galactosyl-(1->3)-N-acetyl-beta-D-glucosaminyl derivative + GDP-beta-L-fucose = a beta-D-galactosyl-(1->3)-[alpha-L-fucosyl-(1->4)]-N-acetyl-beta-

1 Publication

Manual assertion based on experiment (Inferred from experiment)ⁱ

Molecular cloning and characterization of the *Caenorhabditis elegans* alpha1,3-fucosyltransferase family.

[Nguyen K., van Die I., Grundahl K.M., Kawar Z.S., Cummings R.D.](#)

Sezione funzione della voce UniProtKB G5EFP5 (*Caenorhabditis elegans* fut-3) che mostra la carta da cui sono stati estratti i dati.

PROTEINE

- Banche dati proteiche



2. Dati copiati da una proteina caratterizzata sperimentalmente

Per le informazioni che sono state trasferite da una proteina caratterizzata sperimentalmente correlata, viene fornito il numero di accesso della proteina caratterizzata

▶ Region	34-38	Substrate binding	By Similarity
▶ Metal binding	35	Divalent metal cation	By Similarity
▶ Binding site	105	Substrate	By Similarity
▶ Metal binding	196	Divalent metal cation; via carbamate group	By Similarity
▶ Metal binding	227	Divalent metal cation	By Similarity
▶ Metal binding	229	Divalent metal cation	By Similarity
▶ Binding site	362	Substrate	By Similarity

Voce UniProtKB D3DJ41 (*Hydrogenobacter thermophilus* cf*A*) che mostra il numero di accesso della voce da cui è stato trasferito il residuo modificato.

PROTEINE

- Banche dati proteiche



3. Dati importati

Se le informazioni sono state importate da un altro database, vengono forniti il nome del database e l'identificatore della voce da cui sono state importate le informazioni

Gene namesⁱ

Name | Acacb Imported

Synonyms | Acc2 1 Publication, Accb Imported

Organism namesⁱ

Organism | *Mus musculus* (Mouse) Imported

Taxonomic identifierⁱ | 10090 NCBI [↗](#)

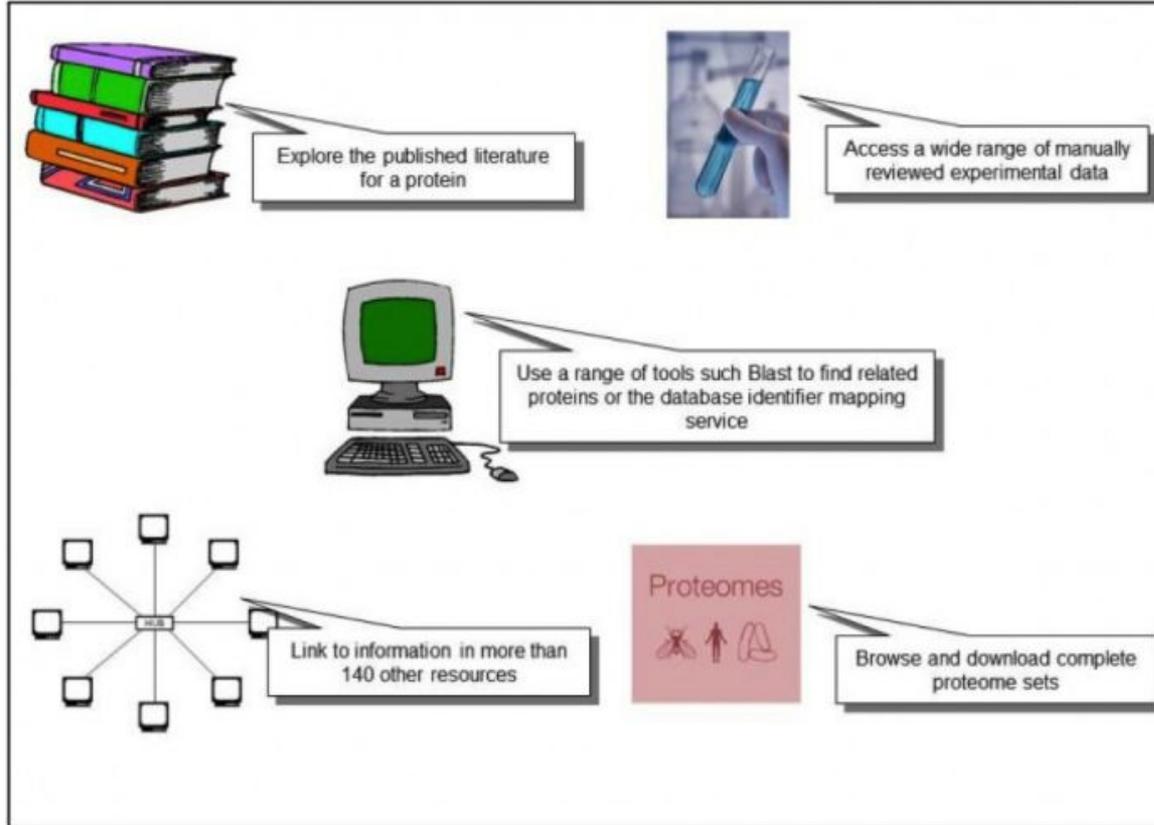
Taxonomic lineageⁱ | > Muscellular organisms > Eukaryota (eucaryotes) > Opisthokonta > Metazoa (metazoans) > Eumetazoa > Bilateria > Deuterostomia > Chordata (chordates) > Craniata > Vertebrata (vertebrates) > Gnathostomata (jawed vertebrates) > Teleostomi > Euteleostomi (bony vertebrates) > Sarcopterygii > Dipnotetrapodomorpha > Tetrapoda (tetrapods) > Amniota (amniotes) > Mammalia (mammals) > Theria > Eutheria (placentals) > Boreoeutheria > Euarchontoglires > Glires (Rodents and rabbits) > Rodentia (rodent) > Myomorpha (mice and others) > Muroidea > Muridae > Murinae > Mus > Mus

Voce UniProtKB E9Q4Z2 (mouse Acacb) che mostra che il nome del gene è stato importato dalla risorsa Mouse Genome Informatics (MGI).

PROTEINE

Quando usare UniProt

È possibile eseguire molte attività diverse utilizzando UniProt, tra cui le seguenti:



PROTEINE

- Banche dati proteiche



Navigazione nei dataset UniProt

UniProt ha una serie di set di dati che puoi navigare e cercare all'interno. È possibile fare clic sul menu a discesa a sinistra della casella di ricerca per visualizzare questi set di dati e selezionare quello che interessa.

Find your protein

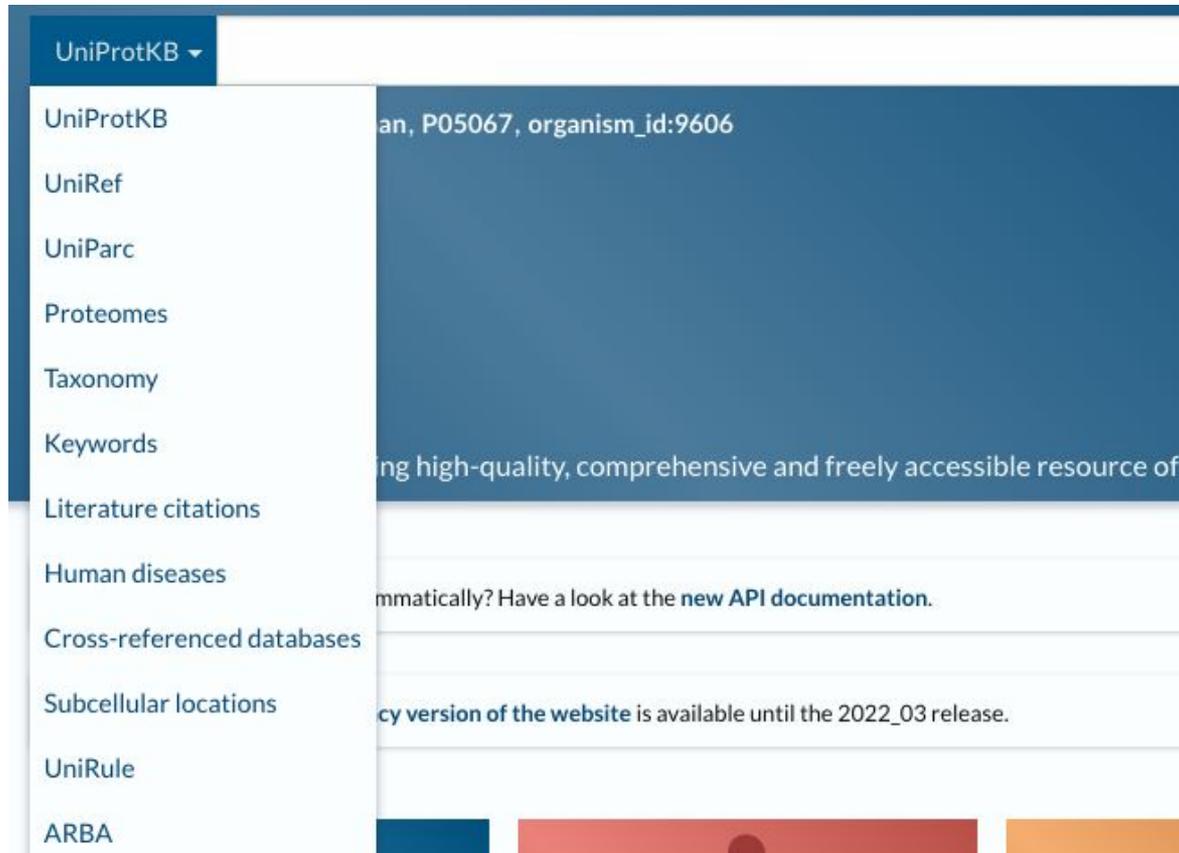
UniProtKB ▾

Advanced | List Search

Examples: Insulin_ABP, Human, P05067, organism_id:9606

PROTEINE

- Banche dati proteiche



The image shows a screenshot of the UniProtKB website. At the top left, there is a dark blue header with the text "UniProtKB" and a small downward arrow. Below this, a white dropdown menu is open, listing several navigation options: UniProtKB, UniRef, UniParc, Proteomes, Taxonomy, Keywords, Literature citations, Human diseases, Cross-referenced databases, Subcellular locations, UniRule, and ARBA. The background of the website is a dark blue gradient with some text and a search bar visible.



Accesso ai dataset UniProt tramite il menu a tendina o riquadri sulla home page.

PROTEINE

- Banche dati proteiche



Navigazione negli strumenti UniProt

UniProt fornisce quattro strumenti principali:

- il Basic Local Alignment Search Tool (**BLAST**) per la ricerca di sequenze
- lo strumento di allineamento di sequenze multiple " **Allinea** "

Analysis Tools

BLAST

Search with a sequence to find homologs through pairwise sequence alignment

Align

Align two or more protein sequences with Clustal Omega to find conserved regions

Search with Lists Map IDs

Find proteins with lists of UniProt IDs or convert from/to other database IDs

Search Peptides

Search with a peptide sequence to find all UniProt proteins that contain exact matches

PROTEINE



- Banche dati proteiche
- lo strumento ' **Recupero/mappatura ID** ' in cui è possibile inviare un elenco di identificatori per recuperare le voci UniProt corrispondenti o mappare gli identificatori UniProt a un database esterno
- lo strumento " **Ricerca peptidi** " che consente di inviare brevi sequenze peptidiche di almeno 3 residui e trovare tutte le sequenze UniProtKB che hanno una corrispondenza esatta con la sequenza query

Analysis Tools

BLAST

Search with a sequence to find homologs through pairwise sequence alignment

Align

Align two or more protein sequences with Clustal Omega to find conserved regions

**Search with Lists
Map IDs**

Find proteins with lists of UniProt IDs or convert from/to other database IDs

Search Peptides

Search with a peptide sequence to find all UniProt proteins that contain exact matches

PROTEINE

- Banche dati proteiche



Visualizzazione dei risultati della ricerca

I risultati di qualsiasi ricerca possono essere visualizzati in (a) Scheda o (b) Vista tabella. La visualizzazione a scheda fornisce un breve riepilogo delle informazioni in una voce, la visualizzazione a tabella è personalizzabile e consente all'utente di decidere quali informazioni aggiuntive desidera visualizzare nei risultati della ricerca.

a UniProtKB 1,530 results or search "Dyrk2" as a Gene Name or Protein Name

BLAST Align Map IDs Download Add View: Cards Table Share

Q92630 · DYRK2_HUMAN
Dual specificity tyrosine-phosphorylation-regulated kinase 2 · Homo sapiens (Human) · EC:2.7.12.1 · Gene: DYRK2 · 601 amino acids · Evidence at protein level · Annotation score: 5/5
#Kinase #Serine/threonine-protein kinase #Transferase #Tyrosine-protein kinase #Apoptosis #Ubl conjugation pathway
1 domain · 6 PTMs · 6 reviewed variants · 1 active site · 2 isoforms · 7 interactions · 22 3D structures · 29 reviewed publications

Q5U4C9 · DYRK2_MOUSE
Dual specificity tyrosine-phosphorylation-regulated kinase 2 · Mus musculus (Mouse) · EC:2.7.12.1 · Gene: Dyrk2 · 599 amino acids · Evidence at transcript level · Annotation score: 6/5
#Kinase #Serine/threonine-protein kinase #Transferase #Tyrosine-protein kinase #Apoptosis #Ubl conjugation pathway

b UniProtKB 1,530 results or search "Dyrk2" as a Gene Name or Protein Name

BLAST Align Map IDs Download Add View: Cards Table Customize columns Share

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	Interacts with
<input type="checkbox"/> Q92630	<input checked="" type="checkbox"/> DYRK2_HUMAN	Dual specificity tyrosine-phosphorylation-regulated kinase 2[...]	DYRK2	Homo sapiens (Human)	601 AA	Q9NR20 (DYRK4) Q13422 (IKZF1) Q9BQD3 (KXD1) Q9BRK4 (LZTS2) P23497 (SP100) 2 more items
<input type="checkbox"/> Q5U4C9	<input checked="" type="checkbox"/> DYRK2_MOUSE	Dual specificity tyrosine-phosphorylation-regulated kinase 2[...]	Dyrk2	Mus musculus (Mouse)	599 AA	

PROTEINE

- Banche dati proteiche



Quando si visualizza una voce UniProtKB, una serie di schede sul lato sinistro dello schermo elenca le opzioni di visualizzazione, consentendo di spostarsi facilmente tra la voce completa, una visualizzazione grafica delle caratteristiche della sequenza (visualizzatore di funzionalità), pubblicazioni, collegamenti esterni ad altre risorse o la cronologia della voce.

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB ▾

★ P11802 · CDK4_HUMAN

Cyclin-dependent kinase 4 · *Homo sapiens* (Human) · EC:2.7.11.22 · Gene: CDK4 · 303 amino acids · Evidence at protein level · Annotation score: 5/5

Entry Feature viewer Publications External links History

BLAST Align [Download](#) [Add](#) Add a publication Entry feedback

Functionⁱ

Ser/Thr-kinase component of cyclin D-CDK4 (DC) complexes that phosphorylate and inhibit members of the retinoblastoma (RB) protein family including RB1 allows dissociation of the transcription factor E2F from the RB/E2F complexes and the subsequent transcription of E2F target genes which are RB1 in early G₁ phase. Cyclin D-CDK4 complexes are major integrators of various mitogenic and antimitogenic signals. Also phosphorylates SMAD3 Component of the ternary complex, cyclin D/CDK4/CDKN1B, required for nuclear translocation and activity of the cyclin D-CDK4 complex. 3 Publications

Catalytic Activity

ATP + L-seryl-[protein] = ADP + H⁺ + O-phospho-L-seryl-[protein] 3 Publications
EC:2.7.11.22 ([UniProtKB](#) | [ENZYME](#) | [Rhea](#))

PROTEINE

- Banche dati proteiche



Le seguenti opzioni di visualizzazione sono fornite per ogni voce UniProtKB.

Iscrizione	Voce completa che mostra tutte le sezioni disponibili
Pubblicazioni	Le pubblicazioni relative alla voce provengono dalla curatela di UniProtKB e quelle mappate computazionalmente da altre fonti
Visualizzatore di funzionalità	Rappresentazione grafica di tutte le caratteristiche della sequenza UniProtKB per l'inserimento in un display interattivo più ulteriori caratteristiche della sequenza da progetti su larga scala
link esterno	Riferimenti incrociati a risorse e pagine Web che forniscono informazioni aggiuntive sulla proteina descritta nella voce
Storia	Una cronologia completa della sequenza e delle annotazioni della voce da quando è apparsa per la prima volta in UniProtKB/TrEMBL

PROTEINE

- Banche dati proteiche



Funzione

Questa sezione di una voce UniProt fornisce un'ampia gamma di informazioni sul ruolo della proteina in una cellula

UniProtKB | BLAST | Align | Peptide search | ID mapping | SPARQL | UniProtKB

P11802 · CDK4_HUMAN

Cyclin-dependent kinase 4 · *Homo sapiens* (Human) · EC:2.7.11.22 · Gene: CDK4 · 303 amino acids · Evidence at protein level · Annotation score: 5/5

Entry | Feature viewer | Publications | External links | History

BLAST | Align | Download | Add | Add a publication | Entry feedback

Functionⁱ

Ser/Thr-kinase component of cyclin D-CDK4 (DC) complexes that phosphorylate and inhibit members of the retinoblastoma (RB) protein family including RB1 allows dissociation of the transcription factor E2F from the RB/E2F complexes and the subsequent transcription of E2F target genes which are RB1 in early G₁ phase. Cyclin D-CDK4 complexes are major integrators of various mitogenic and antimitogenic signals. Also phosphorylates SMAD3 Component of the ternary complex, cyclin D/CDK4/CDKN1B, required for nuclear translocation and activity of the cyclin D-CDK4 complex. 3 Publications

Catalytic Activity

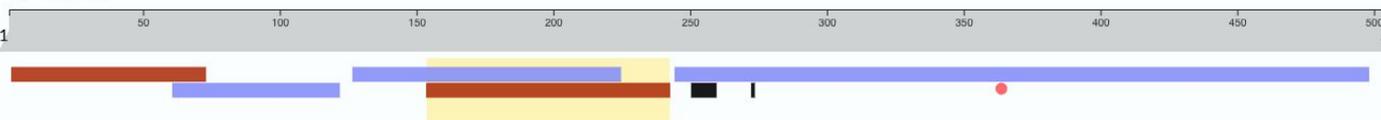
ATP + L-seryl-[protein] = ADP + H⁺ + O-phospho-L-seryl-[protein] 3 Publications
EC:2.7.11.22 (UniProtKB | ENZYME | Rhea)

PROTEINE

- Banche dati proteiche

include una descrizione generale della sua funzione che, nel caso di un enzima, includerà una descrizione della sua attività catalitica, dei cofattori richiesti e potenzialmente anche una dichiarazione di regolamento di attività. Più in generale, verranno descritti i percorsi e i processi in cui è noto che una proteina è coinvolta, nonché altre informazioni correlate come termini di Gene Ontology e siti funzionali come regioni catalitiche o siti di legame all'interno della sequenza proteica, sia graficamente che in un visualizzatore incorporato

Features
Showing features for region¹, domain¹, nucleotide binding¹, binding site¹, active site¹.



TYPE	ID	POSITION(S)	DESCRIPTION	
Region		2-72	Interactions with CD4 and CD8 By Similarity	BLAST
Domain		61-121	SH3 1 Automatic Annotation	BLAST
Domain		127-224	SH2 1 Automatic Annotation	BLAST
Region		154-242	Interaction with PTPRH 1 Publication	BLAST
Domain		245-498	Protein kinase 1 Automatic Annotation	BLAST
Nucleotide binding		251-259	ATP 1 Automatic Annotation	BLAST
Binding site		273	ATP 1 Automatic Annotation	
Active site		364	Proton acceptor 2 Automatic Annotations	

PROTEINE

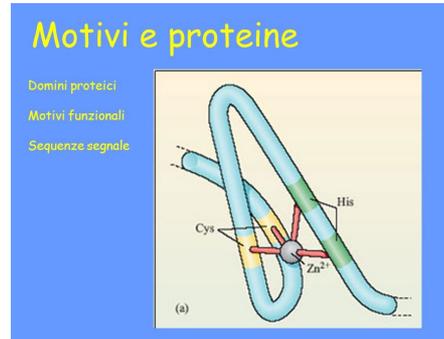
- Banche dati proteiche



Motivi funzionali

L'identificazione dei motivi funzionali proteici può essere utile per **predire la funzione** di una proteina, stabilire **relazioni evolutive** tra proteine di diverse specie e per analizzare e classificare i dati delle analisi proteomiche.

Ma cosa sono i motivi funzionali?



I motivi funzionali proteici, noti anche come domini proteici, sono regioni delle proteine che svolgono una particolare funzione biologica. Essi sono caratterizzati da sequenze di amminoacidi altamente conservate, che si trovano comunemente in proteine di funzione simile in diverse specie. Ad esempio, il dominio SH2 (Src Homology 2) è presente in molte proteine coinvolte nella segnalazione cellulare e contiene un pattern di sequenza conservato di circa 100 amminoacidi.

PROTEINE

- Banche dati proteiche



L'identificazione dei motivi funzionali proteici può essere utile per predire la funzione di una proteina, stabilire relazioni evolutive tra proteine di diverse specie e per analizzare e classificare i dati delle analisi proteomiche.

Tuttavia, l'identificazione dei motivi funzionali proteici può essere problematica in quanto può produrre un gran numero di **falsi positivi**, ovvero sequenze di amminoacidi che presentano somiglianze con la sequenza conservata del motivo funzionale, ma che in realtà non svolgono la funzione specifica del dominio

PROTEINE

- Banche dati proteiche



PROSITE e ELM sono due banche dati di motivi funzionali proteici che aiutano a identificare i motivi funzionali presenti nelle proteine.

PROSITE è una raccolta di motivi funzionali, rappresentati come pattern di sequenza o profili di amminoacidi, accompagnati da informazioni sulle loro funzioni biologiche e sulle proteine che li contengono.

PROSITE offre inoltre la possibilità di effettuare ricerche tramite parole chiave, omologia o sequenze.



PROTEINE

- Banche dati proteiche

ELM (Eukaryotic Linear Motif) è una banca dati di motivi lineari, ovvero brevi sequenze di amminoacidi che svolgono una funzione biologica specifica.



ELM contiene informazioni sui contesti funzionali dei motivi, come l'interazione con altre proteine, la localizzazione cellulare e la regolazione post-traduzionale. Inoltre, ELM permette di cercare i motivi per sequenza o per funzione biologica.

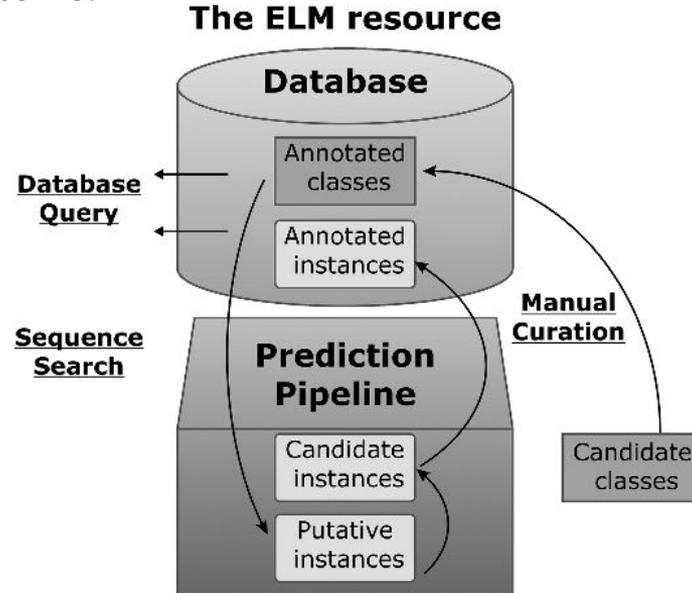


**The Eukaryotic Linear Motif resource for
Functional Sites in Proteins**

PROTEINE

- Banche dati proteiche

Entrambe le banche dati vengono curate da team di esperti e vengono costantemente aggiornate con nuove informazioni e nuovi motivi funzionali scoperti. La loro consultazione può essere molto utile per analizzare e interpretare i dati delle analisi proteomiche e per comprendere meglio le funzioni biologiche delle proteine.



PROTEINE

- Banche dati proteiche



Il PDB contiene informazioni dettagliate sulla struttura tridimensionale di proteine e altri macromolecole biologici determinati mediante cristallografia a raggi X, risonanza magnetica nucleare (NMR) e altri metodi sperimentali.

Le informazioni contenute nel PDB includono le coordinate atomiche dei singoli atomi di proteine e altri macromolecole biologici, nonché informazioni sulla struttura di legame dei ligandi e delle co-fattori.

Atomic Coordinates: PDB Format

	Element	Amino Acid		Chain name		-----Coordinates-----			(etc.)
		Element	Sequence Number	Chain name	Sequence Number	X	Y	Z	
ATOM	1	N	ASP L	1	1	4.060	7.307	5.186	...
ATOM	2	CA	ASP L	1	1	4.042	7.776	6.553	...
ATOM	3	C	ASP L	1	1	2.668	8.426	6.644	...
ATOM	4	O	ASP L	1	1	1.987	8.438	5.606	...
ATOM	5	CB	ASP L	1	1	5.090	8.827	6.797	...
ATOM	6	CG	ASP L	1	1	6.338	8.761	5.929	...
ATOM	7	OD1	ASP L	1	1	6.576	9.758	5.241	...
ATOM	8	OD2	ASP L	1	1	7.065	7.759	5.948	...

\\
Element position within amino acid

PROTEINE

- Banche dati proteiche



Inoltre, ogni entry del PDB include anche informazioni sulle proprietà fisiche e chimiche della molecola, come la sua massa molecolare, il punto isoelettrico, la struttura secondaria e la topologia, nonché informazioni sulla fonte biologica della molecola, come la specie di origine e il tipo di tessuto o cellula in cui è stata trovata. Infine, le entry del PDB possono contenere anche informazioni sulla funzione biologica della molecola, se disponibili.

PROTEINE

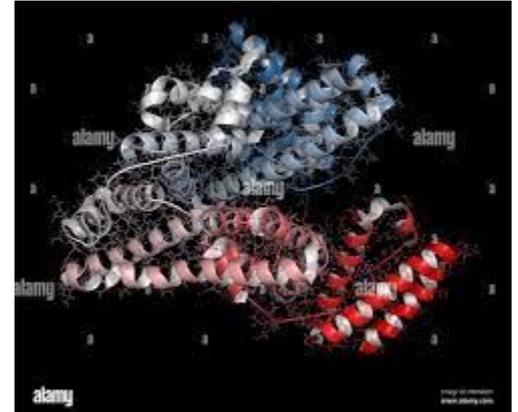
- Banche dati proteiche



I **programmi di grafica molecolare** sono utili nel caso dell'utilizzo del database PDB perché consentono di visualizzare le strutture tridimensionali delle proteine e di interagire con esse in modo intuitivo.

Ciò è particolarmente utile per esplorare la struttura e la funzione delle proteine, per studiare come i legami tra gli atomi di una proteina influenzano la sua struttura e la sua funzione, e per progettare nuovi farmaci o altre molecole che si legano alle proteine.

Quindi i programmi di grafica molecolare sono strumenti essenziali per l'analisi e l'interpretazione dei dati del database PDB e per la progettazione razionale di nuovi composti bioattivi.



PROTEINE

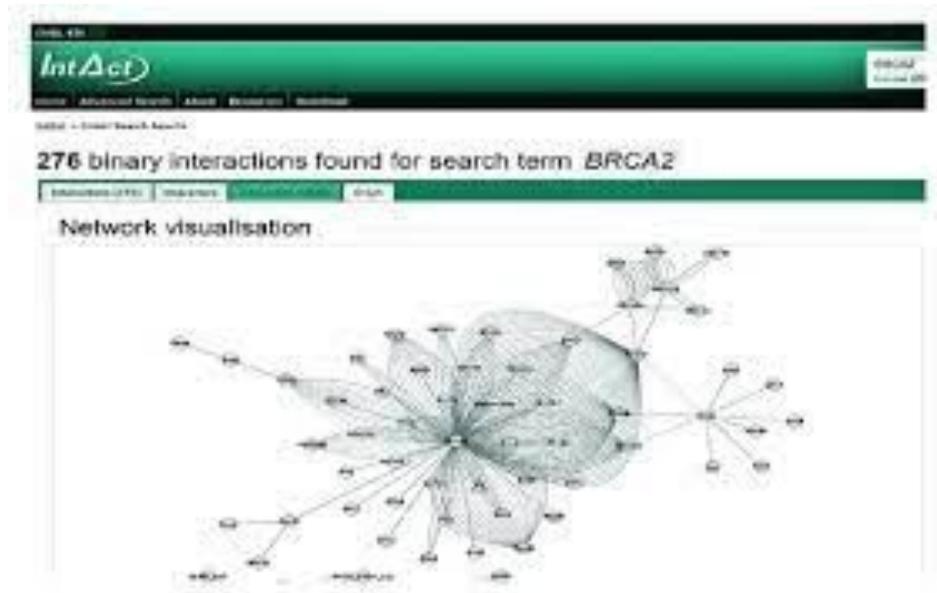
- Banche dati proteiche



Database di interazione proteica:

Un database di interazione proteica è una raccolta di informazioni sulle interazioni tra le proteine. Questi database contengono informazioni su quali proteine interagiscono tra loro, come e quando avviene l'interazione e dove avviene nel contesto cellulare.

L'obiettivo principale di questi database è quello di aiutare i ricercatori a comprendere le reti di interazioni proteiche all'interno delle cellule e la loro relazione con le funzioni cellulari.



PROTEINE

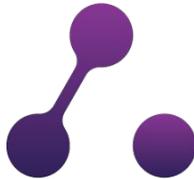
- Banche dati proteiche



Database di interazione proteica:

IntAct (Interaction Database) è un database di interazioni proteiche che raccoglie dati provenienti da letteratura scientifica, da database di pathway e di annotazione funzionale, e da depositi di interazioni proteiche ad alta risoluzione.

IntAct offre un'interfaccia di ricerca user-friendly, che permette di cercare le interazioni tra proteine basandosi su diversi criteri, tra cui il nome della proteina, il gene, la funzione biologica, il sito di interazione, la specie, etc.

IntAct 

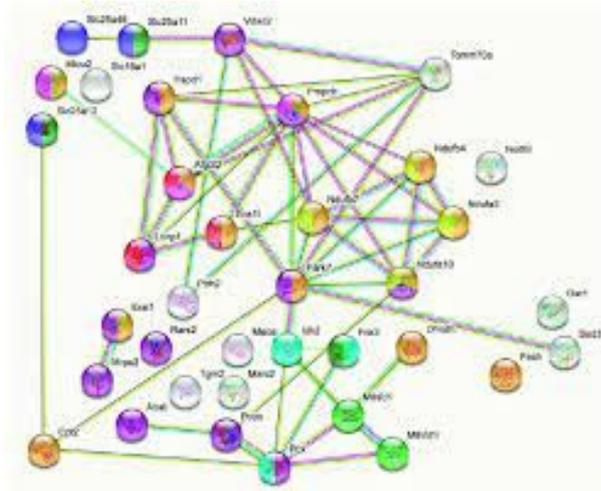
<https://www.ebi.ac.uk/intact/home>

PROTEINE

- Banche dati proteiche



STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) è un database di interazioni proteiche e funzionali. Oltre ad offrire informazioni sulle interazioni tra proteine, STRING integra anche informazioni sulle funzioni biologiche delle proteine e sui pathways di segnalazione in cui queste proteine sono coinvolte. La ricerca di interazioni proteiche in STRING può essere effettuata utilizzando diversi criteri, tra cui il nome della proteina, il gene, la funzione biologica, la specie, etc.

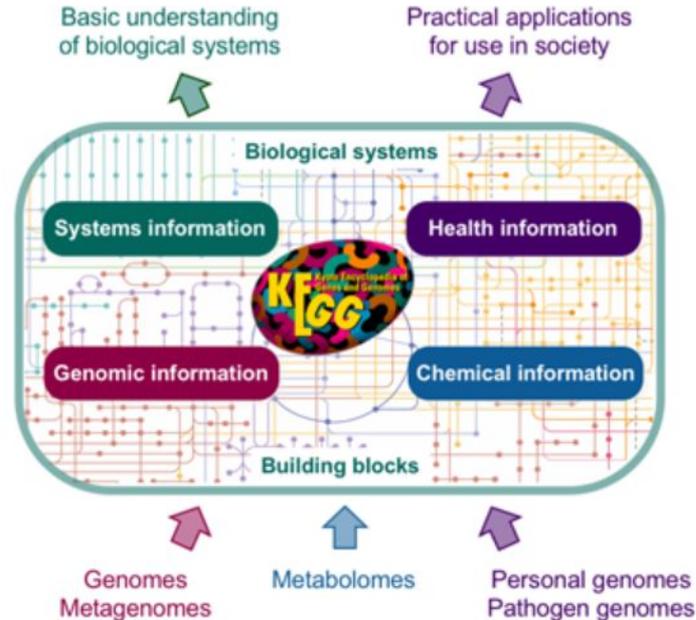


PROTEINE

- Banche dati proteiche



KEGG (Kyoto Encyclopedia of Genes and Genomes) è una banca dati di biologia **sistemica** che integra informazioni genetiche e metabolomiche in una mappa completa di vie metaboliche, vie di segnalazione e malattie associate. KEGG fornisce una raccolta di informazioni sui geni e sui loro prodotti proteici, nonché sulle interazioni tra tali prodotti.



PROTEINE

- Banche dati proteiche



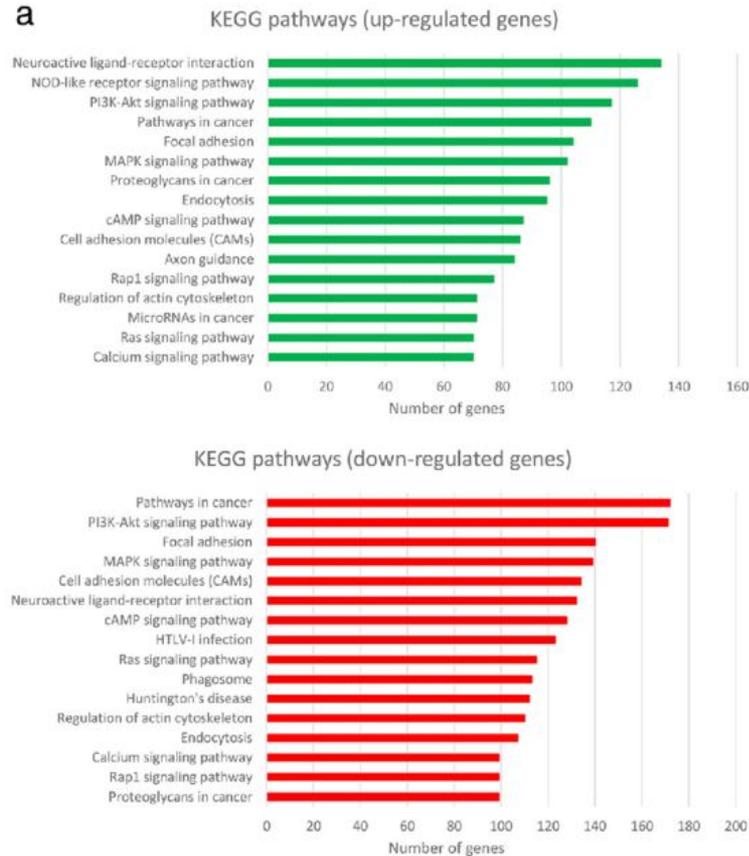
KEGG contiene informazioni genetiche e biochimiche su una vasta gamma di organismi, inclusi batteri, funghi, piante e animali, e include anche informazioni su malattie umane e su come queste sono correlate alle vie metaboliche e di segnalazione. KEGG è organizzato in diversi database, tra cui KEGG Pathway, KEGG Orthology (KO), KEGG BRITE e KEGG Genome.

- KEGG Orthology (KO) è un sistema di classificazione gerarchica per i geni e i loro prodotti proteici. KEGG BRITE è un database di gerarchie di geni e proteine in base alla loro funzione e al loro ruolo all'interno delle vie metaboliche e di segnalazione.
- Infine, KEGG Genome contiene informazioni genomiche su vari organismi, tra cui sequenze di geni, mappe cromosomiche e informazioni sulla struttura genomica.

PROTEINE

- Banche dati proteiche

- KEGG Pathway fornisce una mappa completa di vie metaboliche e di segnalazione, che include informazioni sui geni e sui loro prodotti proteici, nonché sulle interazioni tra tali prodotti.



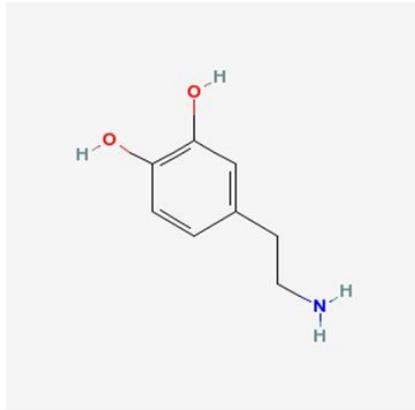
PROTEINE

- Banche dati proteiche

Esempio di output:



Compound Summary:



-  CID: 681 [?](#)
-  **BioActivity: Summary** [?](#)
All: 35 Links
Active: 3 Links
Inactive: 25 Links
Inconclusive: 6 Links
-  **Protein Structures: 2 Links** [?](#)
-  **Protein Sequences: 126 Links** [?](#)
-  **NLM Toxicology:** [?](#)
Link1, Link2
-  **Substances:** [?](#)
All: 63 Links
Same: 48 Links
Mixture: 15 Links
-  **Related Compounds:** [?](#)
Same, Connectivity: 2 Links
-  **Similar Compounds: 72 Links** [?](#)
-  **Structure Search** [?](#)



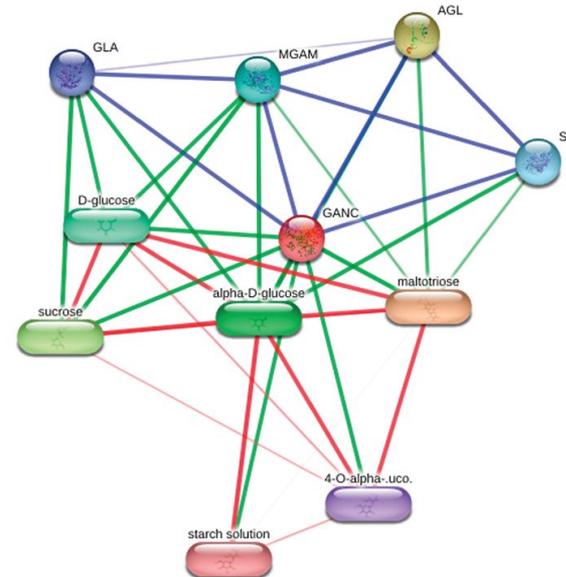
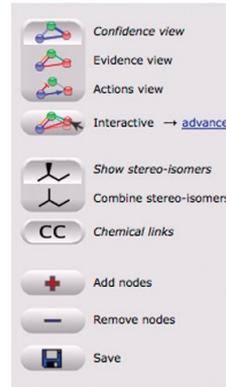
PROTEINE

- Banche dati proteiche

STITCH (Search Tool for Interacting Chemicals) è una banca dati che contiene informazioni sulle interazioni tra composti chimici e proteine.



STITCH integra molteplici fonti di dati, tra cui banche dati proteiche, di interazioni tra proteine e di composti chimici, al fine di fornire una visione globale delle interazioni tra composti e proteine.



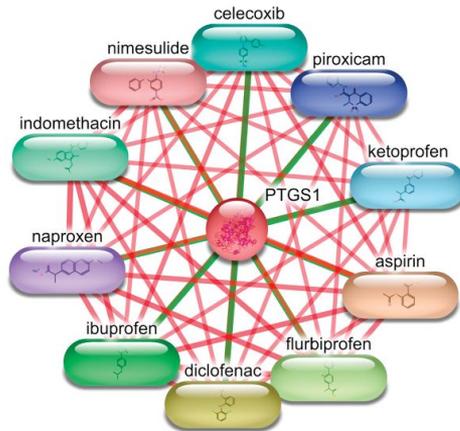
PROTEINE

- Banche dati proteiche



STITCH permette di effettuare ricerche su diversi livelli di risoluzione, dai composti chimici a livello di atomi fino alle interazioni proteina-proteina. Inoltre, STITCH consente di visualizzare le interazioni tra composti chimici e proteine all'interno di pathway metabolici, fornendo una visione integrata delle interazioni tra composti e proteine in contesti biologici specifici.

(a)



(b)

